Supplemental material for "Small area estimation of the proportion of single-person households: Application to the Spanish Household Budget Survey"

María Bugallo Porto[1,*], Domingo Morales González[1] and María Dolores Esteban Lefler[1]

June 2024

The material contained herein is supplementary to the article named in the title and published in SORT-Statistics and Operations Research Transactions Volume 48(1).

[*] *Corresponding author:* mbugallo@umh.es

[1] Operations Research Center, Miguel Hernández University of Elche (Spain). Address: Edificio Torretamarit - Avda. de la Universidad s/n, 03202 Elche (Alicante).

A. ML-Laplace approximation algorithm

This section describes the Laplace approximation of the model log-likelihood and the algorithm to calculate the maximum likelihood (ML) estimators of the model parameters and obtain the modal predictors of the random effects. The function `glmmTMB` of the `R` statistical package `glmmTMB` implements this algorithm. Note that the proposed methodology is based on an area-level zero-inflated Poisson (aZIP13) mixed model (see Section 3). It is therefore a mixture model with Bernoulli (BE) and Poisson (PO) distributions. Although for a mixture-type model it seems more natural to use the EM algorithm, it is not recommended in our research. This is because the EM algorithm does not provide modal predictions of the random effects. Hence, it is a drawback when calculating the IN predictor, as it would have to be obtained using the EBP or ESP predictors of the random effects. In such a case, the IN predictor would no longer have the computational advantages observed in the simulation experiments of Appendix B.

For the sake of completeness, let us start with the Laplace approximation of a multiple integral of a general function $\exp(h(\boldsymbol{x}))$, where $h:\mathbb{R}^m \mapsto \mathbb{R}$ is a twice continuously differentiable function with a global maximum at the column vector $\boldsymbol{x}_0$. That is, let us assume that $\dot{h}(\boldsymbol{x}_0) = \left.\dfrac{dh}{d\boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} = \boldsymbol{0}$ and $\ddot{h}(\boldsymbol{x}_0) = \left.\dfrac{d^2h}{d\boldsymbol{x}^2}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}$ is negative definite.
A Taylor series expansion of $h(\boldsymbol{x})$ around $\boldsymbol{x}_0$ yields to

$$\begin{aligned} h(\boldsymbol{x}) &= h(\boldsymbol{x}_0) + \dot{h}^{\intercal}(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^{\intercal}\ddot{h}(x_0)(\boldsymbol{x}-\boldsymbol{x}_0) + o\big(\|\boldsymbol{x}-\boldsymbol{x}_0\|^2\big) \\ &\approx h(\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^{\intercal}\ddot{h}(x_0)(\boldsymbol{x}-\boldsymbol{x}_0). \end{aligned}$$

It holds that the multivariate Laplace approximation of the integral of $\exp(h(\boldsymbol{x}))$ is

$$\begin{aligned} \int_{\mathbb{R}^m} e^{h(\boldsymbol{x})}\,d\boldsymbol{x} &\approx \int_{\mathbb{R}^m} e^{h(\boldsymbol{x}_0)} \exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^{\intercal}\big(-\ddot{h}(\boldsymbol{x}_0)\big)(\boldsymbol{x}-\boldsymbol{x}_0)\Big\}\,d\boldsymbol{x} \\ &= (2\pi)^{m/2}\ \big|-\ddot{h}(\boldsymbol{x}_0)\big|^{-1/2} e^{h(\boldsymbol{x}_0)}, \end{aligned}$$

where it is used that the integral of the multivariate normal p.d.f. $f(\boldsymbol{x})$ is one.

The likelihood of the aZIP13 mixed model is

$$P(\boldsymbol{y};\boldsymbol{\theta}) = \int_{\mathbb{R}^{K(1+IJ)}} P(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta}) f_u(\boldsymbol{u})\,d\boldsymbol{u} = \int_{\mathbb{R}^{K(1+IJ)}} \exp\big\{h(\boldsymbol{u};\boldsymbol{y},\boldsymbol{\theta})\ \ d\boldsymbol{u}, \qquad \text{(A.1)}$$

where

$$h(\boldsymbol{u};\boldsymbol{y},\boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \log P(y_{ijk}|\boldsymbol{u}_{ijk};\boldsymbol{\theta}) - \frac{K(1+IJ)}{2}\log 2\pi - \frac{1}{2}\sum_{k=1}^{K}\Big(u_{1,k}^2 + \sum_{i=1}^{I}\sum_{j=1}^{J} u_{2,ijk}^2\Big)$$

and $\boldsymbol{u}_{ijk} = (u_{1,k}, u_{2,ijk})^{\intercal}$. To apply the Laplace approximation to the integral in (A.1), we have to maximize $h(\boldsymbol{u};\boldsymbol{y},\boldsymbol{\theta})$ in $\boldsymbol{u}$, given $\boldsymbol{y}$ and $\boldsymbol{\theta}$. For simplicity, we write $h(\boldsymbol{u})$. We

can carry out the maximization by applying a R function of optimization. Alternatively, we can implement a Newton-Raphson algorithm after calculating the first and second partial derivatives of h with respect to $u_{1,k}$ and $u_{2,ijk}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, given $\boldsymbol{y}$ and $\boldsymbol{\theta}$. Let $\dot{h}$ and $\ddot{h}$ denote the $K(1+IJ) \times 1$ vector and the $K(1+IJ) \times K(1+IJ)$ matrix of first and second order partial derivatives of $h(\boldsymbol{u})$ with respect to $\boldsymbol{u}$, respectively. The Newton-Raphson updating equation is

$$\boldsymbol{u}^{(i+1)} = \boldsymbol{u}^{(i)} - \ddot{h}^{-1}(\boldsymbol{u}^{(i)})\, \dot{h}(\boldsymbol{u}^{(i)}). \tag{A.2}$$

Let us denote by $\boldsymbol{u}^{\circ}$ the argument of maxima of the function $h(\boldsymbol{u})$. It holds $\dot{\boldsymbol{h}}(\boldsymbol{u}^{\circ}) = \boldsymbol{0}$ and the matrix $\ddot{h}(\boldsymbol{u}^{\circ})$ is negative definite.

The log-likelihood of the aZIP13 mixed model can be approximated by

$$\log P(\boldsymbol{y};\boldsymbol{\theta},) \approx IJ \log 2\pi + h(\boldsymbol{u}^{\circ}) - \frac{1}{2} \log |-\ddot{h}(\boldsymbol{u}^{\circ})| \triangleq g(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}^{\circ}).$$

The following step is to maximize $g(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}^{\circ})$ in $\boldsymbol{\theta} \in \Theta$. For simplicity, we write $g(\boldsymbol{\theta})$. Maximization can be done by applying a R function of optimization. Alternatively, we can implement a Newton-Raphson algorithm after calculating the first and second partial derivatives of g with respect to the components of $\boldsymbol{\theta}$, given $\boldsymbol{y}$ and $\boldsymbol{u}^{\circ}$. Let be $M = \dim(\Theta) = q_1 + q_2 + 2$. Let $\dot{g}$ and $\ddot{g}$ denote the $M \times 1$ vector and the $M \times M$ matrix of first and second order partial derivatives of $g(\boldsymbol{\theta})$, respectively. The Newton-Raphson updating equation is

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \ddot{g}^{-1}(\boldsymbol{\theta}^{(i)})\, \dot{g}(\boldsymbol{\theta}^{(i)}). \tag{A.3}$$

All things considered, the final ML-Laplace approximation algorithm combines the two described Newton-Raphson algorithms and can be summarized by the following steps:

1. Set the initial values $i = 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\varepsilon_3 > 0$, $\varepsilon_4 > 0$, $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(-1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{1}$, $\boldsymbol{u}^{(0)} = \boldsymbol{0}$, $\boldsymbol{u}^{(-1)} = \boldsymbol{1}$, where $\boldsymbol{0}$ and $\boldsymbol{1}$ are column vectors of zeros and ones, respectively.
2. Until $\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}\|_2 < \varepsilon_1$, $\|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i-1)}\|_2 < \varepsilon_2$, do
   (a) Apply algorithm A.2 with seeds $\boldsymbol{u}^{(i)}$, convergence tolerance ε_3 and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i)}$ fixed. Output: $\boldsymbol{u}^{(i+1)}$.
   (b) Apply algorithm A.3 with seeds $\boldsymbol{\theta}^{(i)}$, convergence tolerance ε_4 and $\boldsymbol{u} = \boldsymbol{u}^{(i+1)}$ fixed. Output: $\boldsymbol{\theta}^{(i+1)}$.
   (c) $i \leftarrow i + 1$.
3. Output: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(i)}$ and $\hat{\boldsymbol{u}} = \boldsymbol{u}^{(i)}$.

The ML-Laplace approximation algorithm is applied to maximize the model log-likelihood (3.2) in the model parameters and in the random effects. In the output, the algorithm approximates the model log-likelihood (which is an integral), calculates the

maximum likelihood estimators of the model parameters and gives predictors of the random effects (mode or modal predictors). The model log-likelihood contains integrals of dimension $1+IJ$.

The ML-Laplace approximation algorithm contains two sub-algorithms (algorithms A.2 and A.3). The first one approximate multiple integrals by applying the Newton-Raphson algorithm (algorithm NR) and by making a maximization on the random effects. The second one performs the algorithm NR to maximize the approximated log-likelihood $h(\boldsymbol{u};\boldsymbol{y},\boldsymbol{\theta})$, given after (A.1), in the model parameters. Algorithm NR search local maxima. However, because of the quadratic form of function $h(\boldsymbol{u};\boldsymbol{y},\boldsymbol{\theta})$, algorithm NR will stop in the neighbourhood of the global maximum. Unfortunately, the second subalgorithm, maximizing $g(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}^\circ)$ in $\boldsymbol{\theta}\in\Theta$, may converge to a local maximum instead of the global maximum. This is why we advice to run the ML-Laplace approximation algorithm starting from appropriate starting values. Our recommendation is to fit a logit mixed model to the zero non-zero data and a Poisson mixed model to the count data, and use the obtained estimates of model parameters as algorithm seeds.

Through the convergence of the ML-Laplace approximation algorithm, besides the ML estimators of the model parameters, it provides modal predictors, $\hat{\boldsymbol{u}}$, of the random effects and the maximized marginal log-likelihood. Since the ML estimators are consistent and asymptotically normal when I, J and K tend to infinity (see e.g. Section 3.7.2 in Jiang (2007)), the algorithm can also be used to approximate the asymptotic covariance matrix (inverse of the Fisher information matrix) which allows the calculation of Wald statistics to test hypotheses about the model parameters. In practice, we use the sign-shifted Hessian matrix (second derivatives of the log-likelihood function) as an approximation of the Fisher information matrix. That is, the asymptotic variance matrix of $\hat{\boldsymbol{\theta}}$, $\boldsymbol{Q}(\boldsymbol{\theta})$, can be approximated as $\boldsymbol{Q}(\boldsymbol{\theta})\approx -\ddot{g}^{-1}(\hat{\boldsymbol{\theta}})$. Further, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is $N_M(\boldsymbol{\theta},\boldsymbol{Q}(\boldsymbol{\theta}))$. Therefore, an asymptotic CI at the level $1-\alpha$ for a component θ_ℓ of $\boldsymbol{\theta}$ is

$$\hat{\theta}_\ell \pm z_{1-\alpha/2}\, q_{\ell\ell}^{1/2},\ \ell=1,\ldots,M,$$

where $\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}^\kappa$, $\boldsymbol{Q}(\boldsymbol{\theta}^\kappa)=(q_{ab})_{a,b=1,\ldots,M}$, κ is the last iteration of the ML-Laplace algorithm and z_α is the α-quantile of the $N(0,1)$ distribution. For a regression parameter $\beta_{a\ell}$, $a=1,2$, $\ell=1,\ldots,q_a$, we can give asymptotic p-values to test significance. For example, if $\hat{\beta}_{1\ell}=\beta_0$, the p-value to test $H_0:\beta_{1\ell}=0$ vs $H_1:\beta_{1\ell}\neq 0$ is

$$\text{p-value}=2P_{H_0}(\hat{\beta}_{1\ell}>|\beta_0|)=2P(N(0,1)>|\beta_0|/\sqrt{q_{\ell\ell}}),\quad \ell=1,\ldots,q_1.$$

To test $H_0:\beta_{2\ell}=0$ vs $H_1:\beta_{2\ell}\neq 0$, we use $q_{q_1+\ell q_1+\ell}$ instead of $q_{\ell\ell}$.

B. Simulations under the SHBS2016 scenario

Based on the case study, i.e. the SHBS2016 data, two simulation experiments have been performed. It should be recalled that domains have been determined according to the $I=52$ Spanish provinces, $J=2$ sexes and $K=4$ age groups. Therefore, there

are $D = IJK = 416$ domains defined by the crosses of provinces, sex and age groups. According to Option 1, the dependent variable y_{ijk} is the direct estimator of the total count of single-person households in province i, with main breadwinner of sex j and age group k. Moreover, we have assumed that y_{ijk} follows the aZIP13 mixed model selected in the statistical analysis of Section 6. Let u_{1k}, $u_{2,ijk}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, be i.i.d. $N(0,1)$ random effects. As $q_1 = 1$, the BE submodel contains one auxiliary variable: $x_{1,1} =$ intercept, with regression parameter $\beta_{11} = -2.696$. The standard deviation parameter is $\phi_1 = 0.398$. Further, the probability parameters of the BE submodel are

$$p_{ijk} \equiv p_k = \exp\{\beta_{11} + \phi_1 u_{1,k}\}\big(1 + \exp\{\beta_{11} + \phi_1 u_{1,k}\}\big)^{-1}, \quad i \in \mathbb{I}, \;\; j \in \mathbb{J}, \;\; k \in \mathbb{K}.$$

The PO submodel contains $q_2 = 4$ auxiliary variables: $x_{2,1} =$ intercept, $x_{2,2} =$ edu3, $x_{2,3} =$ civ2 and $x_{2,4} =$ civ3, with regression parameters $\beta_{21} = -1.857$, $\beta_{22} = 2.138$, $\beta_{23} = -0.649$ and $\beta_{24} = 3.881$. As discussed in Section 6, the remaining variables presented in Section 2 are not incorporated into the model because they are not significant at 5%. The standard deviation parameter is $\phi_2 = 0.5171$. The intensity parameters of the PO submodel are

$$\lambda_{ijk} = \exp\left\{\sum_{\ell=1}^{4} x_{2,ijk\ell}\beta_{2\ell} + \phi_2 u_{2,ijk}\right\}, \quad i \in \mathbb{I}, \;\; j \in \mathbb{J}, \;\; k \in \mathbb{K}.$$

The domain target random quantities are $\mu_{yijk} = m_{ijk}(1 - p_k)\lambda_{ijk}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, where m_{ijk} is the size parameter N_{ijk} of the subpopulation U_{ijk}. Direct estimates of population sizes and area-level auxiliary variables are obtained from the four 2016SLFS. Given their precision, they are considered to be true population values, rather than estimates. Setting the random effects $u_{1,k}$ to their theoretical expected value zero, we compute the basic zero-inflated probability $p_0 = p_0(\beta_{11}) = \exp\{\beta_{11}\}\big(1 + \exp\{\beta_{11}\}\big)^{-1}$, which takes the value $p_0(-2.696) = 0.063$.

B.1. Simulation 1

Simulation 1 aims to assess the fitting algorithm, investigate the performance of the predictors of μ_{yijk} and show how the proposed methodology behaves in simulations compared to other models, in order to identify its advantages. Apart from the predictors derived from the AZIP13 mixed model, i.e., from the SP, ESP and IN, the plug-in predictor with fixed zero-inflated probability (IN1)[1] and the one based on the incorrect non-inflated PO mixed model (IN0)[2] are considered. See Boubeta et al. (2016) for further information about the IN0 predictor.

As far as the other models is concerned, a description of how the framework for comparisons has been set up is given below. Accordingly, and relying on the classical literature of area-level models, the empirical best linear unbiased predictor (EBLUP) of

[1] The parameters of the IN1 predictor are $\beta_{11}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \phi_2$.

[2] The parameters of the IN0 predictor are $\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \phi_2$.

the total of single-person households based on the basic Fay-Herriot (FH) model is included (see Fay, Herriot (1979) for further details). In addition, a zero-inflated negative binomial (NB) mixed model (aZINB13) is also fitted to identify the advantages of the proposed procedure for excess zeros. So, instead of the PO distribution, the count is modelled with a NB distribution. As in Section 3, the random intercept of the BE submodel depends on the age group and that of the NB submodel depends on the domain. The IN predictor is derived. For both models, the same set of auxiliary variables as used in the above-mentioned PO models is considered.

Simulation 1 has the following steps:

1. Repeat $R = 10^3$ times ($r = 1, \ldots, R$):

   1.1. Generate $u_{1,k}^{(r)}$, $u_{2,ijk}^{(r)}$ i.i.d. $N(0,1)$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.

   1.2. For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate

$$p_k^{(r)} = \exp\{\beta_{11} + \phi_1 u_{1,k}^{(r)}\} \left(1 + \exp\{\beta_{11} + \phi_1 u_{1,k}^{(r)}\}\right)^{-1},$$

$$\lambda_{ijk}^{(r)} = \exp\left\{\sum_{\ell=1}^{4} x_{2,ijk\ell}\beta_{2\ell} + \phi_2 u_{2,ijk}^{(r)}\right\}, \quad \mu_{yijk}^{(r)} = m_{ijk}\left(1 - p_k^{(r)}\right)\lambda_{ijk}^{(r)}.$$

   1.3. Generate $z_{ijk}^{(r)} \sim \text{BE}\left(\hat{p}_k^{(r)}\right)$, $y_{ijk}^{(r)} = 0$ if $z_{ijk}^{(r)} = 1$ and $y_{ijk}^{(r)} \sim \text{PO}\left(m_{ijk}\lambda_{ijk}^{(r)}\right)$ if $z_{ijk}^{(r)} = 0$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.

   1.4 For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate $\hat{\tau}^{(r)} \in \{\hat{\beta}_{11}^{(r)}, \hat{\beta}_{21}^{(r)}, \hat{\beta}_{22}^{(r)}, \hat{\beta}_{23}^{(r)}, \hat{\beta}_{24}^{(r)}, \hat{\phi}_1^{(r)}, \hat{\phi}_2^{(r)}\}$ and $\hat{\mu}_{yijk}^{(r)} \in \{\hat{\mu}_{yijk}^{sp(r)}, \hat{\mu}_{yijk}^{esp(r)}, \hat{\mu}_{yijk}^{in(r)}, \hat{\mu}_{yijk}^{in1(r)}, \hat{\mu}_{yijk}^{in0(r)}, \hat{\mu}_{yijk}^{FH(r)}\}$.

2. For each estimator τ and predictor $\hat{\mu}_{yijk}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate

$$BIAS(\hat{\tau}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\tau}^{(r)} - \tau), \quad RMSE(\hat{\tau}) = \left(\frac{1}{R}\sum_{r=1}^{R}(\hat{\tau}^{(r)} - \tau)^2\right)^{1/2},$$

$$BIAS_{ijk} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\mu}_{yijk}^{(r)} - \mu_{yijk}^{(r)}), \quad RMSE_{ijk} = \left(\frac{1}{R}\sum_{r=1}^{R}(\hat{\mu}_{yijk}^{(r)} - \mu_{yijk}^{(r)})^2\right)^{1/2},$$

$$ABIAS = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}|BIAS_{ijk}|, \quad RMSE = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}RMSE_{ijk}.$$

3. Calculate the corresponding relative performance measures in %. That is, calculate the relative bias ($RBIAS_{ijk}$), the relative root-MSE ($RRMSE_{ijk}$), the average absolute relative bias (ARBIAS) and the average relative root-MSE (RRMSE):

$$RBIAS(\hat{\tau}) = 100\frac{BIAS(\hat{\tau})}{|\tau|}, \quad RRMSE(\hat{\tau}) = 100\frac{RMSE(\hat{\tau})}{|\tau|},$$

$$RBIAS_{ijk} = 100\frac{BIAS_{ijk}}{|\overline{\mu}_{yijk}|}, \quad RRMSE_{ijk} = 100\frac{RMSE_{ijk}}{|\overline{\mu}_{yijk}|}, \quad \overline{\mu}_{yijk} = \frac{1}{R}\sum_{r=1}^{R}\mu_{yijk}^{(r)},$$

$$ARBIAS = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}|RBIAS_{ijk}|, \quad RRMSE = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}RRMSE_{ijk}.$$

We first run Simulation 1 by assuming the same model parameters as in the application to real data. To investigate the effect of the basic zero-inflated probability on the performance measures, we also consider the cases $p_0(-1.386) = 0.200$ and $p_0(0) = 0.500$. For the SHB2016 scenario, with $p_0 = 0.063$, Table B.1 presents the results of Simulation 1 for the model parameters. For both BE and PO submodels, the relative biases are small but the RRMSEs are not, implying that the variance is the main component of the MSE. This may be due to the ratio between the number of domains and the number of estimated model parameters, $D/M = 416/7 \approx 60$, which is not large enough to activate the asymptotic properties of the ML estimators. However, it is notable that the RRMSEs of β_{11} and ϕ_2 are particularly good. Appendix C provides additional tables for the corresponding simulation results under scenarios with basic zero-inflated probabilities $p_0 = 0.2$ and $p_0 = 0.5$, allowing us to analyse what happens as p_0 increases.

Table B.1. *Relative performance measures (in %) of the model parameter estimators with* $p_0 = 0.063$*. Simulation scenario based on SHBS2016.*

	BE submodel		PO submodel				
	β_{11}	ϕ_1	β_{21}	β_{22}	β_{23}	β_{24}	ϕ_2
Estimate	-2.696	0.398	-1.857	2.138	-0.649	3.881	0.517
RBIAS	-0.183	-42.196	0.881	-2.303	-1.289	-0.486	-0.510
RRMSE	11.329	78.693	190.866	121.998	326.656	96.880	3.708

Table B.2 provides the relative performance measures of Simulation 1 for the predictors SP, ESP, IN (of the aZIP13 and aZINB13 mixed models), IN1, IN0 and FH. To better understand the necessity of running this experiment and interpret its results, we emphasize that the predictors SP and ESP are not calculated, but rather are approximated, since the integrals that appear in their formulas cannot be calculated analytically. The approximations are obtained by the antithetical Monte Carlo (MC) method, with $S = 2000$, as it is described in Section 4. Since we approximate integrals in $\mathbb{R}^2$, the approximations are not precise enough to acquire the theorical properties. Increasing S even more in a simulation experiment with $R = 1000$ iterations entails unaffordable computation times in Simulation 1 and even more so in Simulation 2. Therefore, the results are subject to the approximation method and the number of iterations.

Table B.2. *Relative performance measures (in %) for the predictors with $S = 2000$. Two model-based alternatives are included. Simulation scenario based on SHBS2016.*

		aZIP13					FH	aZINB13
p_0	Measure	SP	ESP	IN	IN1	IN0	EBLUP	IN
0.063	ARBIAS	0.358	0.361	0.790	9.179	3.068	0.780	3.968
	RRMSE	14.429	14.476	14.662	60.759	60.789	30.380	28.143
0.200	ARBIAS	0.727	0.739	2.444	9.286	13.460	1.405	4.492
	RRMSE	26.270	26.155	25.894	60.714	63.759	57.578	34.117
0.500	ARBIAS	2.965	2.130	5.955	10.716	77.754	2.925	5.566
	RRMSE	43.697	43.075	40.926	62.293	104.149	111.921	44.572

The discussion of Table B.2 starts with the analysis of the predictors proposed for the aZIP13 mixed model. Under all scenarios, the SP has the lowest bias, increasing slightly in its theoretical versions. When substituting true model parameters by ML estimates, the performance of the ESP is almost as good as that of the SP. In fact, changes are minimal. In nominal terms, the variance has a notable contribution to the RMSE for all predictors. Since the ESP and the IN predictor have similar RRMSEs, it has been decided to use the latter in Simulation 2 and in the case study, as its computational cost is lower. As expected, the predictors IN1 and IN0 are biased and have higher RRMSE than the IN predictor. They are based on wrong assumptions.

Under the scenarios with basic zero-inflated probabilities $p_0 = 0.2$ and $p_0 = 0.5$, the predictors IN1 and IN0 perform poorly, with relative biases equal to 9.286 and 13.460 ($p = 0.2$). By increasing p_0 from 0.2 to 0.5, the RRMSE of the IN1 predictor stabilizes and, even though the IN predictor is better, it indicates that age-group randomness is less relevant for such high zero-inflated probabilities. In the latter case, the IN0 predictor performs extremely poorly. To sum up, we conclude that the IN predictor obtained from the aZIP13 mixed model performs much better than the predictor based on the model with constant zero inflation structure. The same applies to the IN0 predictor. Therefore, we do not recommend to use predictors IN0 and IN1 if there is an excess of zeros.

What happens with other solutions based on a model-based approach is discussed below. As for the EBLUP-FH, its bias is small for all values of p_0, with results close to the SP and the ESP. However, this is not a zero-inflated model, which has a negative impact on the error through a significant increase in the variance as p_0 increases. In fact, its RRMSE is even worse than that of IN0 predictor when $p_0 = 0.5$. It has been shown that the response variable has excess zeros and the FH model does not provide a solution to this problem. Regarding the IN predictor of the aZINB13 mixed model, their bias is greater than that of the IN predictor of the aZIP13 mixed model. However, this is compensated to some extent by a lower variance, achieving similar but worse results.

Although both bias and error increase with increasing number of zeros, the proposed ZIP-based estimators perform much better than the traditional FH EBLUP. In fact, ZIP models alleviate this situation and adequately model the excess of zeros. Actually, our contribution has proven to be superior to the FH model and the aZINB13 mixed model. As an overall conclusion, the main advantages of the aZIP13 mixed model over existing

models, and in particular of the IN predictor, are computational performance and reduction in bias and RRMSE. This is why in Section 6 we have focused on analysing the concision between the design-based approach and the model-based approach to SAE.

B.2. Simulation 2

Simulation 2 studies the behaviour of the parametric bootstrap estimator of the MSE of a predictor $\hat{\mu}_{yijk}$ of μ_{yijk}. More concretely, we investigate the behaviour of $mse^*(\hat{\mu}_{yijk})$, which is compared with the empirical MSE of $\hat{\mu}_{yijk}$, obtained from Simulation 1. For illustrative purposes and speed of computation, we select $\hat{\mu}_{yijk} = \hat{\mu}^{in}_{yijk}$. The aim is to give some advice on which B value to choose. The outline is as follows.

1. Take $MSE_{ijk} = RMSE^2_{ijk}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, from Simulation 1.
2. Repeat $R = 500$ times ($r = 1, \ldots, R$):
   - 2.1. As in Simulation 1, generate a sample $(y^{(r)}_{ijk}, x_{1,ijk}, \boldsymbol{x}_{2,ijk})$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.
   - 2.2. Calculate $\hat{\beta}^{(r)}_{11}$, $\hat{\beta}^{(r)}_{21}$, $\hat{\beta}^{(r)}_{22}$, $\hat{\beta}^{(r)}_{23}$, $\hat{\beta}^{(r)}_{24}$, $\hat{\phi}^{(r)}_1$, $\hat{\phi}^{(r)}_2$.
   - 2.3. Repeat B times ($b = 1, \ldots, B$):
     - 2.3.1. Generate $u^{*(rb)}_{1,k}$, $u^{*(rb)}_{2,ijk}$ i.i.d. $N(0,1)$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.
     - 2.3.2. For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate

$$p^{*(rb)}_k = \exp\{\hat{\beta}_{11} + \hat{\phi}_1 u^{*(rb)}_{1,k}\} \left(1 + \exp\{\hat{\beta}_{11} + \hat{\phi}_1 u^{*(rb)}_{1,k}\}\right)^{-1},$$

$$\lambda^{*(rb)}_{ijk} = \exp\left\{\sum_{\ell=1}^{4} x_{2,ijk\ell}\hat{\beta}_{2\ell} + \hat{\phi}_2 u^{*(rb)}_{2,ijk}\right\}, \quad \mu^{*(rb)}_{yijk} = m_{ijk}\left(1 - p^{*(rb)}_k\right)\lambda^{*(rb)}_{ijk}.$$

     - 2.3.3 Generate $z^{*(rb)}_{ijk} \sim \text{BE}\left(\hat{p}^{*(rb)}_k\right)$, $y^{*(rb)}_{ijk} = 0$ if $z^{*(rb)}_{ijk} = 1$ and $y^{*(rb)}_{ijk} \sim \text{PO}\left(m_{ijk}\lambda^{*(rb)}_{ijk}\right)$ if $z^{*(rb)}_{ijk} = 0$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.
     - 2.3.4. Calculate the predictor $\hat{\mu}^{*(rb)}_{yijk}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.
   - 2.4 For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate

$$mse^{*(r)}_{ijk} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mu}^{*(rb)}_{yijk} - \mu^{*(rb)}_{yijk}\right)^2.$$

3. For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate

$$B_{ijk} = \frac{1}{R}\sum_{r=1}^{R}\left(mse^{*(r)}_{ijk} - MSE_{ijk}\right), \quad RE_{ijk} = \left(\frac{1}{R}\sum_{r=1}^{R}\left(mse^{(*r)}_{ijk} - MSE_{ijk}\right)^2\right)^{1/2},$$

$$AB = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}|B_{ijk}|, \quad RE = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}RE_{ijk}.$$

4. Calculate the corresponding relative performance measures in %. That is, calculate the relative bias (RB), the relative root-MSE (RRE), the average absolute relative bias (AARB) and the average relative root-MSE (ARRE):

$$RB_{ijk} = 100\frac{B_{ijk}}{MSE_{ijk}},\ RRE_{ijk} = 100\frac{RE_{ijk}}{MSE_{ijk}},$$

$$ARB = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}|RB_{ijk}|, \quad RRE = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}RRE_{ijk}.$$

The non-relative average measures are not very interpretable because they are conditioned to the large values of the target variable. So, the latter suggests focusing our study on the relative ones. For this reason, Figure B.1 prints five boxplots of RB_{ijk} and RRE_{ijk}, $i \in \mathbb{I},\ j \in \mathbb{J},\ k \in \mathbb{K}$, for $B = 100, 200, 400, 500, 600$.

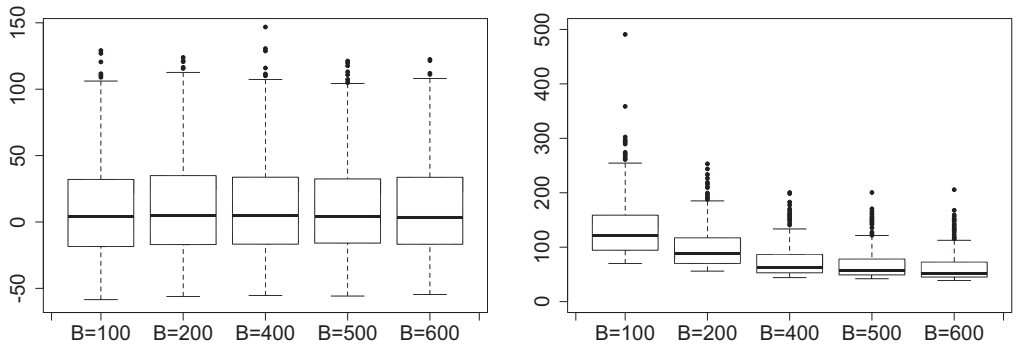

Figure B.1. *Boxplots of RB_{ijk}'s (left) and RRE_{ijk}'s (right), for $B = 100, 200, 400, 500, 600$.*

Table B.3. *Average relative performance measures for $B = 100, 200, 400, 500, 600$.*

B	100	200	400	500	600
ARB	10.244	10.566	10.657	10.723	10.594
RRE	134.643	99.255	73.821	68.054	60.089

As can be observed in Figure B.1 (left), the relative biases do not decrease as the size of B increases, showing an origin-centric behaviour. There are few atypical values that correspond to the most conflictive domains, i.e., those with smaller sample sizes or with zero observed single-person households in SHBS2016. This severely distorts the symmetry of the ordinate axis. On the other hand, Figure B.1 (right) shows that the relative root-MSEs decrease as B increases. Table B.3 confirms this behaviour, with an ARB stabilized around 10 and a RRE decreasing as B increases, but suggesting some stabilization around $B = 600$ iterations. It is concluded that the results for the MSE estimator of the IN predictor are reasonable in most domains. However, the low sample

size of some of them and the non-observation of single-person households increases its bias and, therefore, the error of the parametric bootstrap estimator of the MSE.

C. Additional simulation results

This section provides additional results of Simulation 1 for zero-inflated probabilities $p_0 = 0.2$ and $p_0 = 0.5$. Tables C.1 and C.2 presents the relative performance measures of the ML model parameter estimators. It can be noticed that the estimators of the BE submodel parameters, β_{11} and ϕ_1, have slightly lower values of RBIAS and RRMSE than the corresponding ones under the SHBS2016 scenario, with $p_0 = 0.063$. This suggests that the estimators of the BE submodel perform slightly better if the basic zero-inflated probability increases. However, the changes are minor. For the remaining coefficientes, there are no remarkable differences. It can be argued that the performance of the fitting algorithm is not expected to worsen if the basic zero-inflated probability increases drastically (from 0.063 to 0.2 or even to 0.5).

Table C.1. *Relative performance measures of model parameter estimators with $p = 0.2$. Simulation scenario based on SHBS2016.*

	BE submodel		PO submodel				
	β_{11}	ϕ_1	β_{21}	β_{22}	β_{23}	β_{24}	ϕ_2
Estimate	-2.696	0.398	-1.857	2.138	-0.649	3.881	0.517
RBIAS	0.791	-34.480	1.325	-3.198	-2.113	-0.801	-0.659
RRMSE	16.398	60.556	190.449	122.655	345.917	96.770	4.007

Table C.2. *Relative performance measures of model parameter estimators with $p = 0.5$. Simulation scenario based on SHBS2016.*

	BE submodel		PO submodel				
	β_{11}	ϕ_1	β_{21}	β_{22}	β_{23}	β_{24}	ϕ_2
Estimate	-2.696	0.398	-1.857	2.138	-0.649	3.881	0.517
RBIAS	NaN	-29.43	1.576	-3.116	-3.407	-0.731 3	-1.110
RRMSE	NaN	55.028	191.085	123.530	315.871	96.855	4.994

D. RRMSE maps for the IN predictor of the proportion of single-person households by domains

This section maps the RRMSE of the IN predictor of the proportion of single-person households by domains, estimated by parametric bootstrap with $B = 1000$ resamples. For futher details, see Section 5. Recall that domains are defined as crosses between

provinces, sex and age groups. The maps on which RRMSE estimates are reported are included in Section 6.3 for the SHBS2016 data.

Figures D.1-D.4 show the results for men (left) and women (right), by age group of the main breadwinner, from top to bottom. To sum up, it can be seen that the error varies with province, sex and age group. In fact, as a relative measure, it tends to be higher in those domains where the IN proportions are lower. Overall, it follows that the accuracy of our results is statistically reasonable, with RRMSEs below 30% in most domains, exceeding it only in those where predicted proportions are tiny.

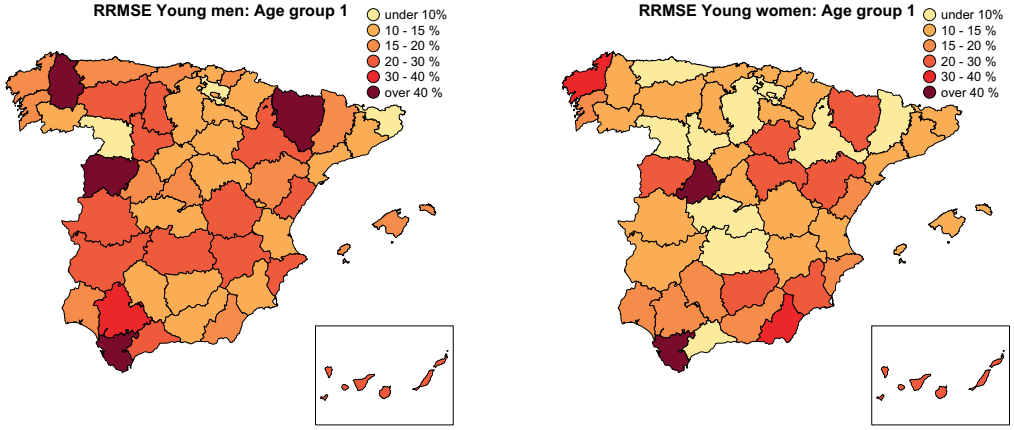

Figure D.1. *RRMSE of the IN predictor of the proportion of single-person households for young men (left) and women (right). Data from SHBS2016.*

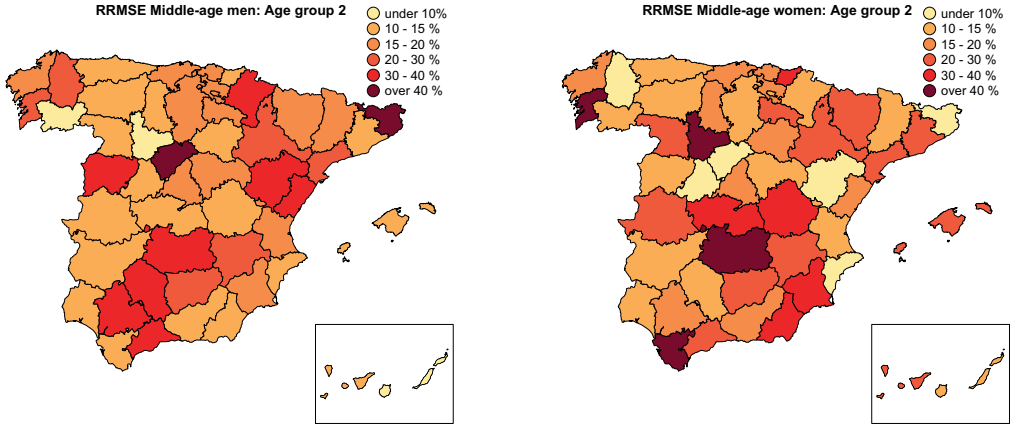

Figure D.2. *RRMSE of the IN predictor of the proportion of single-person households for middle-age men (left) and women (right). Data from SHBS2016.*

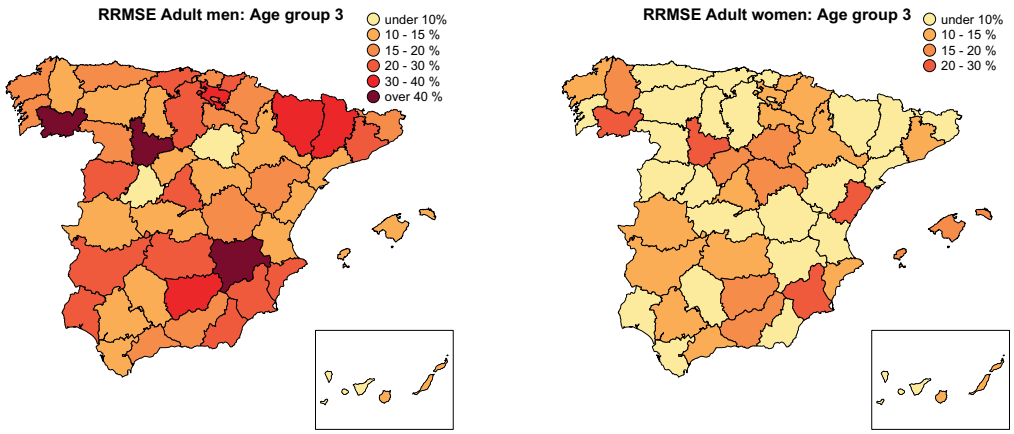

Figure D.3. *RRMSE of the IN predictor of the proportion of single-person households for adult men (left) and women (right). Data from SHBS2016.*

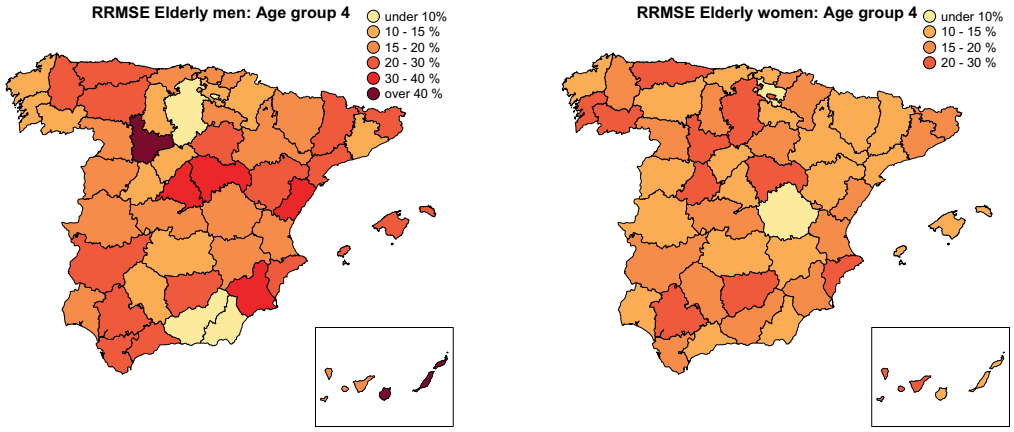

Figure D.4. *RRMSE of the IN predictor of the proportion of single-person households for elderly men (left) and women (right). Data from SHBS2016.*

References

Boubeta, M., Lombardía, M.J. and Morales, D. (2016). Empirical best prediction under area-level Poisson mixed models. *TEST*, 25, 548-569.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*. 74, 269-277.